NPS-AS-92-021

# NAVAL POSTGRADUATE SCHOOL
## Monterey, California



**A Proposal to Apply Taguchi-Inspired Methods to the Reduction of Machining Variance**

**by**

**Dan Trietsch**

**September 1992**



Approved for public release; distribution is unlimited.

Prepared for: Naval Postgraduate School
Monterey, Ca 93943

**NAVAL POSTGRADUATE SCHOOL**
**Monterey, California**

RADM. R. W. West, Jr.
Superintendent

Harrison Shull
Provost

This report was prepared for and funded by the Naval Postgraduate School.



This report was prepared by:

Dan Trietsch

Dan Trietsch
Department of Administrative Sciences

Reviewed by:

David R. Whipple, Chairman
Department of Administrative Sciences

Released by:

Paul J. Marto, Dean of Research

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. AGENCY USE ONLY *(Leave Blank)* | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 16 September 1992 | Technical Report, 1992 |

**4. TITLE AND SUBTITLE**
A Proposal to Apply Taguchi-Inspired Methods to the Reduction of Machining Variance

**5. FUNDING NUMBERS**
O&MN

**6. AUTHOR(S)**
Dan Trietsch

**7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)**
Administrative Sciences Department
Naval Postgraduate School, Monterey, CA 92943

**8. PERFORMING ORGANIZATION REPORT NUMBER**
NPS-AS-92-021

**9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Naval Postgraduate School
Monterey, CA 92943

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release; Distribution Unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

The Probability a Machined part will be defective increases with the variance of the machined dimensions. Even for parts within tolerance, the quality decreases with the variance. By reducing the variance of these dimensions better parts will be produced. Several factors, some of which are controllable, impact this variance, and they may also interact with each other. By choosing an appropriate level for each controllable factor we can minimize the variance. Since factors may interact with each other, a factorial experimental design is appropriate to optimize the levels of the factors; optimizing one factor at a time is not likely to yield the global optimum in the presence of interactions. The designed experiments must be conducted in stages: (i) at the preliminary stage one determines which factors and interactions are likely to be important; (ii) at the exploratory stage one runs a fractional design to identify the factors that are really important and to verify/update the preliminary hypotheses about the interactions; (iii) at the search stage one uses a sequence of experiments with varying levels of the factors that were identified as important to optimize their levels; finally, (CONTINUE AT NEXT PAGE)

| 14. SUBJECT TERMS | 15. NUMBER OF PAGES |
|---|---|
| | 35 |
| | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

Block 13

(iv) at the verification stage one runs the process under the levels deemed to be optimal, to check if the process behaves according to the predictions. At any stage, after analyzing the current data, one may have to go back to a previous stage. This report presents a list of factors and interactions (preliminary stage output), and a design for the exploratory stage. We also discuss how to interpret the results and conduct the search stage. The plan is designed as a generic blueprint that can be used at any machine shop.

NPS-AS-92-021

# A Proposal to Apply Taguchi-Inspired Methods to the Reduction of Machining Variance

by

Dan Trietsch*

September 1992



* Code AS/Tr, Naval Postgraduate School, Monterey, CA 93943

Acknowledgement: The author wishes to thank Professor Max Woods for useful discussions relating to the proposed use of Fisher's z statistic for variance reduction.

# A Proposal to Apply Taguchi-Inspired Methods to the Reduction of Machining Variance


## *Abstract*

The probability a machined part will be defective increases with the variance of the machined dimensions. Even for parts within tolerance, the quality decreases with the variance. By reducing the variance of these dimensions better parts will be produced. Several factors, some of which are controllable, impact this variance, and they may also interact with each other. By choosing an appropriate level for each controllable factor we can minimize the variance. Since factors may interact with each other, a factorial experimental design is appropriate to optimize the levels of the factors; optimizing one factor at a time is not likely to yield the global optimum in the presence of interactions. The designed experiments must be conducted in stages: (i) at the preliminary stage one determines which factors and interactions are likely to be important; (ii) at the exploratory stage one runs a fractional design to identify the factors that are really important and to verify/update the preliminary hypotheses about the interactions; (iii) at the search stage one uses a sequence of experiments with varying levels of the factors that were identified as important to optimize their levels; finally, (iv) at the verification stage one runs the process under the levels deemed to be optimal, to check if the process behaves according to the predictions. At any stage, after analyzing the current data, one may have to go back to a previous stage. This report presents a list of factors and interactions (preliminary stage output), and a design for the exploratory stage. We also discuss how to interpret the results and conduct the search stage. The plan is designed as a generic blueprint that can be used at any machine shop.

## Introduction

The appellation "Taguchi Methods" is a registered trademark denoting a set of off-line quality improvement methods promoted by Genichi Taguchi [Taguchi, 1986]. Chief among them is the use of statistical experimental methods adapted by Taguchi for the purpose of minimizing the "Taguchi loss function," which is really the well-established quadratic loss function that statisticians have been using for decades.

Traditional statisticians treated the variance as a given, and minimized the loss function by adjusting the mean to the required dimension. In contrast, Taguchi places equal stress on reducing the variance. In fact, for the quadratic loss function the expected loss is the sum of the squares of the bias and the standard deviation; i.e.,

$$E(L) = (M - \mu)^2 + \sigma^2$$

where L is the loss, M is the target, $\mu$ is the mean of the distribution and $\sigma$ is its standard deviation. Thus, reducing the bias (hitting the target on average) is not more important then reducing the standard deviation.

Prominent western statisticians fault Taguchi on technical details of his applications, including his use of signal-to-noise ratios that are supposed to make possible the reduction of bias and variance at the same time and the unnecessary confounding of main effects with interactions in his recommended designs [Box Bisgard & Fung, 1988]. They also say, however, that Taguchi's objective and his generic idea to use statistical experimental design to achieve the objective are highly positive contributions that led to remarkable successes already. Thus, what we want to do is to pursue Taguchi's objective (of reducing the variance as well as the bias), using statistical experimental design methods, but not the ones

recommended by Taguchi. We call the resulting framework *Taguchi-Inspired Methods*. This proposal is about applying Taguchi-inspired methods to machining parts. The statistical design will be based on traditional western methods, as described in Box, Hunter and Hunter [1978], but with added modules for the reduction of variance. The reader is referred to that source for technical details that are not covered here, such as: generators, aliases, confounding and defining contrasts.

This plan is designed to reduce the number of defective parts in machine shops and improve the quality of good parts by achieving machined dimensions that are distributed closer to the nominal (target). In machining it is not difficult to adjust the mean, so here we concentrate on reducing the variance (though, for completeness, we'll also provide some details that relate to the adjustment of the mean). This will have a positive effect on the quality of parts, and on meeting tolerances in particular.

Generally, the number of defective parts should be reduced by optimally controlling all the factors involved. These include long-term factors (i.e., factors that can only be changed in the long term), medium-term factors, and short-term factors. The design of the machined part is an example of a long-term factor; e.g., optimizing tolerances based on the use of the part is an immediate way in which the design impacts the number of defective items [Taguchi, 1986]. The maintenance level of machines is an example of a medium-term factor; i.e., once a machine creates losses that exceed the cost of restoring it to its designed operating characteristics, it should be fixed (this generally happens long before the machine actually breaks down, so it's an issue of preventive maintenance rather than corrective maintenance). Changing the width and depth of the machined chips, as well as controlling

the machining speed and the coolant type are examples of short-term factors that can be changed from one part to the next.

Here we limit ourselves to short-term factors only. Many medium- and long-term factors should be addressed only when the short-term operating procedure is optimized. For instance, to determine the losses involved in not fixing a particular machine we need to know that we are doing the best we can with it, otherwise we may be overstating the losses.

Thus, the assumption is that the parts are designed with given tolerances from given materials, and that this design cannot be changed. Under this assumption, the probability a machined part will be acceptable increases when the variance of the machined dimensions is reduced, because this makes it more likely that the tolerances will be met.

Several short-term factors, some of which are controllable, impact this variance, and they may also interact with each other. By choosing an appropriate level for each controllable factor we can minimize the variance.

Since factors may interact with each other, a factorial experimental design, which changes the levels of many factors at the same time, is appropriate to optimize the levels of the factors. In contrast, optimizing one factor at a time is not likely to yield the global optimum in the presence of interactions [Box, Hunter & Hunter, 1978].

Such an experiment must be conducted in stages: (i) at the preliminary stage one determines which factors and interactions are likely to be important; (ii) at the exploratory stage one runs a fractional factorial design to identify the factors that are really important and verify/update the preliminary hypotheses about the interactions; (iii) at the search stage one uses a sequence of experiments with varying levels of the factors that were identified

as important to optimize their levels; finally, (iv) at the verification stage one runs the process under the levels deemed to be optimal, to check if the process behaves according to the predictions.

Ideally, the preliminary stage should be done in cooperation with the machinists who own the process. This plan, however, is designed to give decision makers a flavor of what resources a statistical experiment design might require, and thus has to be written *before* the preliminary stage can take place. Therefore, the author had to resort to his own knowledge of machining to substitute for the preliminary stage. As a result, it is possible that the actual factors involved in a real experimental design may differ from the ones presented here. Henceforth we'll assume that the results of the preliminary stage are as follows.

## List of Factors and Likely Interactions

The following factors were deemed likely to be significant:

1) Depth of chip [D]
2) Width of chip [W]
3) Speed of cut [S]
4) Tool selection (material, make, etc) [T]
5) Coolant fluid [C]
6) Machine selection [Mc]
7) Material of machined part [Mt]

Note that the last two factors are not controllable, but they still introduce variance to the system. Therefore they have to be included in the design. It is possible, however, to conduct the experimentation for a specific machine, in which case [Mc] drops out. Technically, it just means that we can ignore it, as well as any pair that includes it.

The following interactions were deemed likely to be significant, in descending order:

1) [D][W]

2) [D][S]

3) [W][S]

4) [W][T]

5) [D][T]

6) [S][T]

7) [T][C]

8) [S][Mc]

This list will be useful for the design of the exploratory experiments.

## The Exploratory Stage

At the exploratory stage we conduct the first experiments to check out our tentative hypotheses on how the machining process behaves. These hypotheses, in turn, are the output of the preliminary stage. It is important not to spend too much resources at the exploratory stage, since it is quite likely that after analyzing the results the design will have to be changed. (In fact, at any stage, after analyzing the current data, one may have to go back to a previous stage.)

Also, even if the factors identified at the preliminary stage are the pertinent ones, it is quite likely that the levels at which one sets them at the exploratory stage are far from optimal (because the optimum is unknown). Therefore, additional runs will almost surely be required. As a rule of thumb, the exploratory stage should be designed to consume up to 25% of the total experimentation budget.

Since it's not likely that the level of the factors set at the exploratory stage is close to optimal, it is often enough at this stage to only use two levels for each factor. If possible, the use of more levels should be restricted to the search stage (where we use a "star design" for this purpose, which enhances the basic two-level design). In this section, we limit ourselves to a design with two levels per factor.

In general, such fractional factorial designs have $2^3 = 8$, $2^4 = 16$, $2^5 = 32$, $2^6 = 64$ and higher powers of 2 runs each, and each run may also be replicated. In our case, 8 runs are not likely to be enough for our needs, so a fractional design with 16 runs is the least we can do at the exploratory stage. Since it is important not to spend too much resources at this stage, we look for a design with 16 runs, and propose to run it without replicate runs.

With our seven factors, if we wish to limit ourselves to 16 runs, we have to use a fraction of 1/8 (a full factorial with seven factors requires $2^7 = 128$ runs). Thus a $2^{7-3}$ fractional factorial design is suggested. It is relatively easy to replace one of the factors suggested by any new factor. It is even possible to extend the number of factors to eight, as we'll discuss later. Thus the fact that the preliminary stage was not carried out properly yet is not likely to cause major changes in the proposed design--it will probably require 16 runs.

Our original list included seven factors and eight hypothesized interactions. The most economical $2^{7-3}$ design available requires 16 runs without replications and is capable of computing seven main factor effects and seven two-factor interactions. The other 14 theoretical interactions that may exist are confounded (or aliased) with these seven interactions (i.e., they cannot be told apart based on the numerical results alone). Therefore, if we insist on all eight interactions we'll have to run a larger fraction. To prevent this, the two least likely interactions are designed to be confounded together, while the other six interactions are confounded with interactions that are not expected to be significant at all.

For the first three factors, [D], [W] and [S] (depth, width and speed) it is proposed to use 67% and 133% of the values deemed best by the experience of the operators (note that the high level will thus be twice the low level). To limit the number of levels of each factor to two in the last four factors it is proposed to (i) use the best available tool as the high [T] level and the cheapest tool as the low [T] level; (ii) use a similar scheme for [C]; (iii) use the best machine and the worst machine for the high and low levels of [Mc]; and, finally, (iv) use steel and bronze for the high and low levels of [Mt].

*The statistical design*

There is only one resolution IV fractional factorial $2^{7-3}$ design. Let the seven factors be A, B, C, D, E, F, G, then only A through D belong to the full $2^4$ factorial design at the root of the fractional design. Factors E through G have to be generated from A through D. Appropriate generators are: E = ABC; F = BCD; G = ACD. (Theoretically, we could replace any of these generators by ABD, or add an eighth factor with this generator--as

discussed below. Nevertheless, if we use ABD instead of one of the other generators we'll obtain a permutation of the same design. Since we did not assign factors to the letters A through G yet, all these permutations can be obtained by the present choice. Hence the claim that this is the only design that needs to be considered.)

The defining contrast of the design is {I, ABCE, BCDF, ACDG, ADEF, BDEG, ABFG, CEFG}, and we see that this is indeed a resolution IV design (all the expressions in the defining contrast, except for I, have at least four letters).

Using the defining contrast we obtain the following triplets of two-factor interactions that are confounded with each other:

(A) BD; CF; EG

(B) AD; CG; EF

(C) AF; BG; DE

(D) AB; CE; FG

(E) AG; BF; CD

(F) AC; BE; DG

(G) AE; BC; DF

By observing the list we see that each two-factor interaction is confounded with two other two-factor interactions, thus forming seven equivalence groups of three two-factor interactions each. We also see that if a factor appears in any interaction in such a group, it does not appear in any other interaction in the same group. In fact, since we have seven factors and exactly six of them are involved in each equivalence group, then each such group has exactly one factor missing; indeed the seven equivalence groups correspond to the seven

factors: each of them has a different factor missing. These missing factors can thus be used to identify the equivalence groups, and this was done in the list above (e.g., the first line, (A), lists three two-factor interactions that do not include the factor A).

As a result of the observations above, it becomes clear that two interactions involving the same factor cannot be confounded with each other under this design (e.g., [W][S] cannot be confounded with [D][S] because they both have [S] in common). It is also clear that at most we can have seven two-factor interactions that are not confounded with each other. This means we must accept that two of our eight two-factor interactions will have to be confounded. Furthermore, we cannot choose any two such interactions because the two interactions cannot share any factor. Observing the list of interactions above, we come to the conclusion that the potential confounding can only occur between the following pairs of interactions:

1) [D][W] and [S][T]

2) [D][W] and [T][C]

3) [D][W] and [S][Mc]

4) [D][S] and [W][T]

5) [D][S] and [T][C]

6) [W][S] and [D][T]

7) [W][S] and [T][C]

8) [W][T] and [S][Mc]

9) [D][T] and [S][Mc]

10) [T][C] and [S][Mc]

Compared to the theoretical limit of 28 this list is small, but it still leaves the door open to several designs. To choose between the ten possible confounding scheme we observe that (10) involves the pairs we judged to be least likely to be significant (which does not mean that the factors involved are not likely to be significant--just the interactions). Thus we look for a design that confounds [T][C] and [S][Mc], and assigns all the other two-factor interactions to unique equivalence groups.

Fortunately, such a design exists. To obtain it we assign the letters A through G of the design listed above to the actual factors as follows:

1) Depth of chip [D], assigned to B
2) Width of chip [W], assigned to C
3) Speed of cut [S], assigned to A
4) Tool selection (material, make, etc) [T], assigned to D
5) Coolant fluid [C], assigned to E
6) Machine selection [Mc], assigned to F
7) Material of machined part [Mt], assigned to G

The assignment order is almost identical to the order of the factors, but this is a chance occurrence. (There exists another assignation that confounds [T][C] and [S][Mc], which yields a slightly different order.) The following table lists the aliasing involved in this assignation:

(A) **[D][T]**; [W][Mc]; [C][Mt]

(B) **[S][T]**; [W][Mt]; [C][Mc]

(C) **[S][Mc]**; **[T][C]**; [D][Mt]

(D) **[D][S]**; [W][C]; [Mc][Mt]

(E) **[W][T]**; [S][Mt]; [D][Mc]

(F) **[W][S]**; [D][C]; [T][Mt]

(G) **[D][W]**; [S][C]; [T][Mc]

The interactions in bold print are the ones listed as likely to be important. For instance, according to the first line, the interaction between depth and tool selection, [D][T], which is one of our list and therefore bold, is aliased with the interaction between width and machine selection and the interaction between coolant selection and the material, both of which are not expected to be significant. The third line shows that the interactions between speed and machine selection and between tool selection and coolant selection are aliased (as we intended), along with the interaction between depth and material. It's easy to verify that all the other important interactions are assigned to unique equivalence groups, as required.

Table 1 lists this design in Yates order (i.e., the order in which a $2^k$ factorial is constructed). A minus (-) sign in the table signifies low level, and a plus (+) indicates high level. Since a factorial experiment has to be run at random order, the first column is a computer generated order.

| Random order | Setting | [S] | [D] | [W] | [T] | [C] | [Mc] | [Mt] |
|---|---|---|---|---|---|---|---|---|
| 2 | (1) | - | - | - | - | - | - | - |
| 7 | $scm_t$ | + | - | - | - | + | - | + |
| 12 | $dcm_c$ | - | + | - | - | + | + | - |
| 3 | $sdm_cm_t$ | + | + | - | - | - | + | + |
| 8 | $wcm_cm_t$ | - | - | + | - | + | + | + |
| 14 | $swm_c$ | + | - | + | - | - | + | - |
| 6 | $dwm_t$ | - | + | + | - | - | - | + |
| 10 | sdwc | + | + | + | - | + | - | - |
| 9 | $tm_cm_t$ | - | - | - | + | - | + | + |
| 4 | $stcm_c$ | + | - | - | + | + | + | - |
| 1 | $dtcm_t$ | - | + | - | + | + | - | + |
| 13 | sdt | + | + | - | + | - | - | - |
| 5 | wtc | - | - | + | + | + | - | - |
| 16 | $swtm_t$ | + | - | + | + | - | - | + |
| 11 | $dwtm_c$ | - | + | + | + | - | + | - |
| 15 | $sdwtcm_cm_t$ | + | + | + | + | + | + | + |

**TABLE 1**

The settings column is a shorthand notation for the information that follows it in the row. A lower case letter, such as s, indicates that the factor of the same letter, such as S, is at the high level, or +; the lack of such a letter indicates low level, or -. The first row has all levels set low, and therefore the convention (1) is used as shorthand notation. The last row, all pluses, has all the letters present. Since [Mc] and [Mt] use two letters each, a subscript was introduced to differentiate between them, thus $m_c$ indicates that [Mc] is to be

at the high level. For instance, the line before last is marked by $dwtm_c$, which indicates that depth, width, tool and machine have to be set at the high level, while speed, coolant and material have to be set at a low level.

To calculate main effects of factors we take the sum of the results of runs for which the column of the factor is (+), subtract from it the sum of the other runs, for which the sign is (-), and divide the result by 8. A similar calculation applies to the calculation of effects of interactions; here the sign should be taken as the product of the signs of the individual factors involved. For instance, to calculate the effect of the interaction [S][D], we take the first run, where [S] and [D] are both (+), as (+), while the second run, with (-) and (+) will be signed (-), etc.

Note that this calculation method simply takes the average of the difference between runs where the factor (or interaction) is at a high level, denoted by (+), and runs where it is at a low level, (-). For 16 runs there are always eight pluses and eight minuses for each effect and each interaction.

Table 2 repeats the same design, but in the prescribed random order that should be followed in practice, as indicated in the first column of Table 1. Thus the first run, $dtcm_t$ is set with the depth, tool, coolant and material at their high levels and the width, speed and machine at their low levels. The second run sets all factors at their low levels, etc.

| Run # | Setting | [S] | [D] | [W] | [T] | [C] | [Mc] | [Mt] |
|---|---|---|---|---|---|---|---|---|
| 1 | $dtcm_t$ | - | + | - | + | + | - | + |
| 2 | (1) | - | - | - | - | - | - | - |
| 3 | $sdm_cm_t$ | + | + | - | - | - | + | + |
| 4 | $stcm_c$ | + | - | - | + | + | + | - |
| 5 | wtc | - | - | + | + | + | - | - |
| 6 | $dwm_t$ | - | + | + | - | - | - | + |
| 7 | $scm_t$ | + | - | - | - | + | - | + |
| 8 | $wcm_cm_t$ | - | - | + | - | + | + | + |
| 9 | $tm_cm_t$ | - | - | - | + | - | + | + |
| 10 | sdwc | + | + | + | - | + | - | - |
| 11 | $dwtm_c$ | - | + | + | + | - | + | - |
| 12 | $dcm_c$ | - | + | - | - | + | + | - |
| 13 | sdt | + | + | - | + | - | - | - |
| 14 | $swm_c$ | + | - | + | - | - | + | - |
| 15 | $sdwtcm_cm_t$ | + | + | + | + | + | + | + |
| 16 | $swtm_t$ | + | - | + | + | - | - | + |

**TABLE 2**

***Potential design changes***

As mentioned above, it's possible to change the identity of any factor in the list above. The design, however, is geared to the hypotheses about the important interactions. Therefore some potential changes may not be as efficient in terms of the number of interactions that can be investigated efficiently.

Furthermore, we can even add an eighth factor to the list, and as long as it is not expected to have important interactions with the existing factors the design will still work. With seven factors we could identify seven main effects and seven interactions, or a total of 14. The number of degrees of freedom available, after using one for the grand mean, is 15, which implies that one more interaction can be identified, but it would have to be a complex interaction involving more than two factors. With eight factors we can use the same 15 degrees of freedom to identify eight main effects and, again, seven interactions. This explains how we can add the eighth factor without increasing the number of runs.

If we add an eighth factor to the design, say H (using A through G for the original 7 factors), the new generator is H = ABD (and, as before, E = ABC, F = BCD and G = ACD). The defining contrast, which now has 16 elements instead of 8, is: {I, ABCE, BCDF, ACDG, ADEF, BDEG, ABFG, CEFG, ABDH, CDEH, ACFH, AEGH, BEFH, BCGH, DFGH, ABCDEFGH}, where the first eight elements are as before, and all the new elements include H. Analyzing which two-factor interactions will be aliased now, we obtain the following list:

(A) **AH**; BD; CF; EG

(B) AD; **BH**; CG; EF

(C) AF; BG; **CH**; DE

(D) AB; CE; **DH**; FG

(E) AG; BF; CD; **EH**

(F) AC; BE; DG; **FH**

(G) AE; BC; DF; **GH**

The interactions in bold print are the new ones, and, not surprisingly, they each involve the interaction of the factor that was missing in the line before with the new factor, H.

Designating the new factor as [X], for the unknown new factor to be included, and assuming the original seven factors will remain intact, we obtain a new design, presented with the same random running order as before:

| Run # | Setting | [S] | [D] | [W] | [T] | [C] | [Mc] | [Mt] | [X] |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $dtcm_t$ | - | + | - | + | + | - | + | - |
| 2 | (1) | - | - | - | - | - | - | - | - |
| 3 | $sdm_cm_t$ | + | + | - | - | - | + | + | - |
| 4 | $stcm_c$ | + | - | - | + | + | + | - | - |
| 5 | wtcx | - | - | + | + | + | - | - | + |
| 6 | $dwm_tx$ | - | + | + | - | - | - | + | + |
| 7 | $scm_tx$ | + | - | - | - | + | - | + | + |
| 8 | $wcm_cm_tx$ | - | - | + | - | + | + | + | + |
| 9 | $tm_cm_tx$ | - | - | - | + | - | + | + | + |
| 10 | sdwc | + | + | + | - | + | - | - | - |
| 11 | $dwtm_c$ | - | + | + | + | - | + | - | - |
| 12 | $dcm_c$ | - | + | - | - | + | + | - | - |
| 13 | sdtx | + | + | - | + | - | - | - | + |
| 14 | $swm_cx$ | + | - | + | - | - | + | - | + |
| 15 | $sdwtcm_cm_tx$ | + | + | + | + | + | + | + | + |
| 16 | $swtm_t$ | + | - | + | + | - | - | + | - |

**TABLE 3**

*Analyzing the results of the exploratory stage*

To analyze the results of the experiment, the first step is to identify the significant factors and interactions. The fact that our design is fractional may complicate this task, however. For instance, we can never be one hundred percent sure whether two significant effects that are confounded with each other do not cancel each other out. More importantly, we can never be one hundred percent sure which of the aliased effects is the significant one. Here we suggest the use of prior knowledge to make that call, but there is a risk involved. Telling aliased effects apart, when necessary, can be resolved by additional experimentation. Furthermore, it may be necessary to specify transformations of the data to obtain the best representation of the behavior of the model.

Be that as it may, traditional experimental design techniques cover these issues adequately, so we do not dwell on them further here. Instead, let us concentrate on identifying factors and interactions that impact the variance: our main concern. There are two major approaches to the problem, depending on which measurements we take for each run.

One possibility is to try to assess the variation of each part as measured in different locations. The assumption is that the variation will manifest on each part. For instance, it is not likely that the diameter of a lathed part will be absolutely identical at different locations along the shaft and when measured at different radial angles. If this variation is representative of the general variation a particular setting imparts to the processed part, we can obtain a measure of variation, $S^2$, for each run. In this case, it is customary to take the logarithm of $S^2$ as our measure, and we proceed to look for significant effects by one of the traditional methods.

The recommended way to identify such significant interactions is by using the Daniel method. It involves plotting the computed effects on a normal probability paper, and checking which of them, if any, does not lie on a straight line through the central values. Such outliers, if they are in the direction away from the line, are deemed significant. There is a subjective judgement involved, but very often the call seems clear.

Traditionally, this method is applied to the determination of significant effects in terms of impacting the mean response, but there is no reason not to apply it to our logarithms: their distribution tends to be approximately normal.

There is another method, appropriate for the exploratory stage, which requires only one measurement per run. It is especially useful for cases where each run cannot yield information about variance. We may apply it if we have reason to doubt that the variation along the part is indicative of the variation between parts. We discuss how to apply this method below, and, since it is not well-documented in the literature, we justify its use in the Appendix.

Our first step is to find the significant effects in terms of impacting the mean response (i.e., measurement of machined part), as is traditionally done (i.e., by the Daniel method). Next, we measure the residuals for the 16 runs after taking into account these significant effects (or we could use regression analysis for the same purpose). Now, take any factor or interaction we wish to check for, say i, and eight of the runs have it with a plus sign, and eight with a minus sign. Denote the sum of the squares of the eight residuals with the plus sign as $s^2(i^+)$, and, similarly $s^2(i^-)$ denotes the sum of squares of the other residuals. Using these two values we can calculate a combined statistic for each factor or interaction: $\ln[s^2(i^+)/s^2(i^-)]$. If the factor has no impact on the variance, then this statistic has the z distribution (but multiplied by 2), which is Fisher's (earlier) version of Snedecor's F statistic. In fact, Snedecor developed the F tables at Fisher's suggestion for the convenience of people

who lacked tables of the natural logarithm [Fisher, 1990]. (For some unspecified reason Fisher did not want to use logarithms to the base of 10, though it would merely require a multiplication by $\log_{10}e$, i.e., 0.43429, to overcome the lack of these tables. In the opinion of this author, it's most unfortunate that the F format, which Fisher himself thus launched, is so much more prevalent than the z format. The z format is better for interpolation, according to Fisher, and even more importantly, it tends to normal much faster than does the F, either for cases--like ours--when the number of degrees of freedom is equal, or for cases where that number is large. In fact, for the equal number of degrees of freedom case the z approximates the normal better than student's t with the sum of degrees of freedom).

As Fisher notes, the z statistics behaves approximately as a normal random variable when the number of degrees of freedom is equal (but not a standard normal; the mean is zero but the standard deviation is not one), though its tails may be thicker when few degrees of freedom are involved. Therefore, we can apply the Daniel method, again, this time to the 16 values we obtain, and identify outliers (i.e., points that are off the straight line) as we did for the main effects. Any such outliers represent factors or interactions that impact the variance, positively or negatively (depending on which side they are on). This provides us with preliminary information how to proceed.

Another potentially fruitful approach to identify significant effects in terms of variation is the use of interaction plots. Sometimes they reveal that at a particular level of a factor the variation of the output with other factors is diminished: i.e., at this level the factor helps to control the variance.

Note that the last two methods are suggested for the preliminary analysis of the exploratory stage only. During the ensuing search we have to either use replication or use the first method to come up with logarithms of $S^2$ statistics. The reason for this is that at the search stage it is more important to use exact methods.

## Response Surface Methods for Finding Improved Settings: The Search Stage

This section gives a very cursory presentation of the theory of response surface analysis, which is recommended for the search stage. An excellent source for this purpose is Box and Draper [1987]. Though they do not treat the case of minimization of variance specifically, the mathematical methods are the same.

After running the exploratory experiment, and performing some screening and sorting, we screen out the factors that are inert both in terms of mean-effect and variance-effect, and sort the other factors to two (not necessarily exclusive) groups: Group 1 includes effects that impact the mean, while Group 2 includes factors that impact the variance (or at least, we did not rule out the possibility that they do).

Note that when some factors are attribute factors (e.g., black or white, with or without), our task is simply to pick the better level; but if the variables are real (continuous). the optimum may reside in values such as 2.11, rather than exactly at one of the levels we started with. We concentrate on real variables, henceforth.

In general, we may have to look for the best settings both in terms of obtaining the correct average response, and minimizing the variance. For our specific example, machining, there is no problem in adjusting the mean, so we can concentrate on minimizing the variance. In this section, however, we still allow for the possibility of having to adjust the mean too. Nevertheless, for simplicity, we proceed under the assumption that the two groups are disjoint, which also implies that no significant interactions exist between factors across groups. Otherwise, the mathematical task ahead of us is more complex, because we'll have to solve simultaneously for more variables. If the groups are not disjoint we may also have to settle on a point with a bias, because our objective is to minimize $(\sigma^2+\text{bias}^2)$.[1] If the

[1]Actually, our real objective is to minimize the total loss involved, i.e., $k(\sigma^2+\text{bias}^2)+C$, where C is the cost imparted by the setting we choose. This will prevent choosing excessively expensive solutions that cost more than they save. We will not do that here for

groups are disjoint, however, we achieve this end by searching for the best values for the factors in Group 2 in terms of minimizing $\sigma^2$, and, conceptually separately, the best values for the factors in Group 1 to minimize the absolute value of the bias.

We said "conceptually separately" but in reality we can run new settings for the two searches at the same time. For instance, suppose the best guesstimate for the setting in which we expect to achieve the smallest variance is by reducing C by 3 units, and the best guess for the point that minimizes the bias is (A=1.2, B=2.4), then we may wish to test, say, four points in the direction of the vector (A=1.2, B=2.4, C=-3).

Such points may be (0.6, 1.2, -1.5),[2] (1.2, 2.4, -3), (1.8, 3.6, -4.5), and (2.4, 4.8, -6). They include the expected best point as point #2, and the third and fourth points are actually over (to make sure that there are no further gains to be had in this direction). At each point we may decide to take several replicates, to get a good estimate for both the bias and the variance.

It is important to note that when we take replicate runs at the same setting there are two alternatives we can consider. We can randomize the sequence of all the replicates, or we can specify that the replicates be taken consecutively. We recommend the latter.

For instance, suppose that we want to investigate four points, and take five replicates at each point. We should randomize the sequence of the settings, but run the 5 replicates consecutively for each setting, rather than randomize the 20 runs. The advantage of this scheme, in addition to convenience, is that it provides a blocking effect. The variation within

---

two reasons: (i) we may not know k; (ii) for simplicity in presentation, this being an introductory treatise.

[2]The length (or norm) of this first vector is $\sqrt{(0.6^2+1.2^2+1.5^2)}=2.01$; this is not far from the present design, whose radius (in coded variable distances) is $\sqrt{3}=1.73$. The idea is to have the first point within the general neighborhood of the space covered by the present grid, and then venture out in steps of approximately the same size.

each group of replicates is due to the setting itself, and not due to variation in the ambient conditions that may occur between blocks or due to adjustment differences among the replicates. Since our objective is to minimize the variance caused by the setting, it would be detrimental to allow variations between settings to enter the picture.

After running these four points we will search for the best value along the line for C independently from the search for A and B. That is, we'll minimize the bias and the variance independently. As for A and B, we'll only consider A and B combinations that are on the line: to do anything else would be to ignore the possible interaction between them. Thus we may conclude that our best position is, say, (1.5, 3, -2), where C is between the first and second points while A and B are between the second and third points.

Suppose we search in the direction of steepest improvement from the origin, and determine that as long as we're constrained to be on this line, the best value is at or near (1.5, 3, -2). We now run a few (2 to 5, say) replicates at this point (which we did not try before, remember, both in terms of A and B and in terms of C, and certainly not when combined).

These runs are used to make sure that the results we have are as expected, but, more importantly, they'll serve as center points for a new factorial that we now run at this stage. The idea is to either find a new improvement direction where we'll repeat the same process, or do some fine tuning by a quadratic approximation.

Some background that we did not cover so far is in order. For illustration we'll refer to the optimum setting as a mountain top. The idea is that we wish to climb to the top of the mountain, but visibility is poor, so we must base our search on isolated data at the points we already experimented with. We use the mountain top analogy because it has a positive connotation, and also because one cannot expect to gravitate there without conscious effort; but note that in our case we look for the least variance: the lowest point.

The mathematical problem, of course, is practically identical.

Often, when we're really far from the mountain top, a planar approximation describes the response surface quite well. This implies that the interaction effects, which are quadratic in nature, tend to be small relatively to the main effects. But, when we approach the mountaintop, curvature becomes evident, and in such cases the interactions tend to become strong relatively to the main effects.

Another evidence for such curvature would be that the average of the runs on the vertices of the cube (or square or hyper-cube), i.e., the runs where all the settings are at $\pm 1$, is significantly different from the average of the center runs. The difference, by the way, estimates the sum of the coefficients of the squares of the main factors.

When we have such evidence, we know that we're at the general area of the mountaintop, and we can start planning our final runs. The idea is to provide data for a quadratic regression model of the form

$$\textit{Response} = b_0 + \Sigma_i b_i x_i + \Sigma_i \Sigma_j b_{ij} x_i x_j$$

where $b_0$ is the grand mean, $b_i$ is the main effect of $x_i$, $b_{ij}$ for $i \neq j$ denotes two-factor interactions, and $b_{ii}$ is the coefficient for the pure quadratic effect obtained by squaring $x_i$.

But note that we already have several runs in the neighborhood, which we used to verify that the quadratic terms have indeed become important. We also have some central points (which we also used for that purpose). Still, we need more points to be able to estimate the pure quadratic effects, because most of our variables are at the $\pm 1$ levels. Had they been all $\pm 1$, then all quadratic variables would be confounded with the grand mean (taking the form of I, the identity vector). As it is, the central points prevent this mathematically, but the regression would still be of low quality. To mend this we now add what's known as *star* points, of the form $(\pm\alpha, 0, ..., 0)$, $(0, \pm\alpha, 0, ..., 0)$, ...., $(0, ..., 0, \pm\alpha)$.

The star points may be replicated, depending on the noise, the importance of hitting the top closely, and the budget. In our case, since we want to estimate the variance as well as the mean, it's highly recommended to run some replicates. We also add some more central points to the star points, to be able to tell if there are blocking effects between the initial "cube" runs and the subsequent star runs.

Choosing the exact values of $\alpha$ is the subject of complex theory, with conflicting objectives to boot. A good choice that achieves one of these objectives (orthogonality) exactly, and is not far from the values that achieve the other (rotatability) is also simple. Therefore we may choose to adopt it as a rule. The idea is to set the distance of the star points from the center equal to that of the cube points. With k factors this implies $\alpha = \sqrt{k}$. For our example we'll have (1.73, 0, 0), (-1.73, 0, 0), (0, 1.73, 0), (0, -1.73, 0), (0, 0, 1.73), and (0, 0, -1.73).

It remains to discuss two similar mathematical constructs that we need to actually implement the procedure described above. One is to estimate the directions of steepest improvement, and the other is to estimate the exact settings associated with the top of the mountain. For the former we have identified in the initial experiment(s) real effects and, potentially, some interactions. For the latter we have the results of a quadratic regression model run on the points of the cube *plus* the points of the star, and along with the center points from both series (potentially we may have to adjust the values if the block influence is large; we ignore this issue here, however).

For finding the direction of steepest improvement, we'll discuss in particular the two factors case. Our factors are denoted by x and y, and based on the preliminary runs we have the main effects 2a and 2b, respectively. The constant, 2, reflects the fact that the main effects measure the average difference between -1 and +1, i.e., 2 units. In the first case we also have an interaction, 2c. Look at the regression model

$$Response = ax + cxy + by$$

Note that x and y are measured in scaled units based on our choices at the preliminary runs. For instance, suppose x ranged from 1.2 at the minus level to 1.5 at the plus level (and thus x=0 at 1.35), and y ranged from 2 to 3 respectively (with y=0 at 2.5). In this case we can rewrite the regression model in terms of the original units, as follows:

$$Response = \frac{2a(x-1.35)}{1.5-1.2} + \frac{4c(x-1.35)(y-2.5)}{(1.5-1.2)(3.0-2.0)} + \frac{2b(y-2.5)}{3.0-2.0}$$

The most convenient procedure, however, is to stick to the coded variables, and convert the final search results to the original units at the end.

If our response here is of the type smaller-is-better (as is always the case with the variance reduction search), or larger the better, then we need to minimize or maximize the expression, and it's likely that our solution will be unique. If we have to hit a target, however, we often have discretional choices how to set the variables. In such cases the best approach might be to treat the equation as a constraint, and minimize some cost function associated with the variables' values.

Mathematical programming is the tool of choice for such a task, but a practical solution might simply be to set the variables that are associated with high costs to their cheapest levels (or most profitable levels), and adjust the response to the target by one of the variables that are not cost-sensitive.

To minimize or maximize the response, the key is to set the derivative of the response function to zero. But, since the function has several variables, we need to set all the partial derivatives, also known collectively as the gradient, to zero. A partial derivative, recall, is obtained for a variable, say x, by considering all the other variables as constants while taking the derivative by x. When we do that we obtain the following set of linear equations.

$$cy = -a$$

$$cx = -b$$

And it follows that y=-a/c, and x=-b/c. Note (for future reference) that x/y=b/a.

Now, this point that we just calculated is, *by extrapolation*, either a minimum point or a maximum point. We don't know which, but it should be easy to tell at a glance if the point is in the direction of ascent or descent. If we seek a minimum and the point is an extrapolated candidate maximum location, we'll probably want to search in the exact opposite direction. Another potential difficulty is that the point may be too far out; actually, if c is small relatively to $\sqrt{(a^2+b^2)}$ the point *will* be far out. This means that we cannot trust it. It's exactly analogous to making decisions about the form and size of the whole mountain based on a few points at the mountain's foot.

So, we may want to limit the distance we go to, say, a scaled distance of twice the distance from the origin to the grid points. This would translate to a constraint of the form $x^2+y^2 \leq 8$. With such a constraint, even if c=0, the point we'll find will be in the direction of the vector (b, a) or in the opposite direction, depending on whether we want to increase or decrease the response. That is, it will be in the same direction we computed above for the unconstrained optimum when c is not zero.

We now discuss the mathematical method of estimating where the top of the mountain is, based on the quadratic regression model. In fact, this is done by the exact same method as above: setting the gradient of the function to zero. Since the function is quadratic, to set the gradient to zero we have to solve a set of linear equations, which we present in terms of three factors (the generalization is immediate)

$$2b_{11}x_1 + b_{12}x_2 + b_{13}x_3 = -b_1$$

$$b_{21}x_1 + 2b_{22}x_2 + b_{23}x_3 = -b_2$$

$$b_{31}x_1 + b_{32}x_2 + 2b_{33}x_3 = -b_3$$

The solution, to be reliable, should fall within the space covered by the design. Otherwise, we may have to repeat the process by covering the space around the solution with new runs. Finally, even if the solution is comfortably near, the next and last stage involves performing confirming runs at the recommended solution to check if the results fit the predictions.

## References


Box George E.P., Soren Bisgard and Conrad Fung (1988), An Explanation and Critique of Taguchi's Contribution to Quality Engineering, *Quality and Reliability Engineering International*, Vol. 4.

Box, George E.P., William G. Hunter and J. Stuart Hunter (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley.

Box George E.P., and Norman R. Draper (1987), *Empirical Model Building and Response Surfaces*, Wiley.

Fisher, Sir Ronald A. (1990), *Statistical Methods, Experimental Design, and Scientific Inference*, Oxford Science Publications, Oxford (a reprint of three books by the late Sir Ronald A. Fisher).

Taguchi, Genichi (1986), *Introduction to Quality Engineering*, Asian Productivity Organization, Tokyo.

## Appendix

In this appendix we discuss our usage of the statistic $\ln[s^2(i^+)/s^2(i^-)]$ to identify significant effects for the purpose of impacting the variance. First, we have to make an assumption, which is not very strong for large experiments: we assume that the residuals (after adjusting for the real effects) are distributed normally. Given that assumption, there is a powerful traditional tool we can use: the F distribution. The F distribution was developed in the thirties by Snedecor, at Fisher's suggestion [Fisher, 1990]. Fisher's original version, the z statistic, had the format of $\ln[s^2(i^+)/s^2(i^-)]/2$, which, except for the division by 2, is identical to our statistic. The F distribution is concerned directly with the ratio of variances, without bothering with the log transformation. Let's elaborate this point a bit.

Suppose we want to test whether a particular factor, or even an interaction, influences the variance. The basic idea is to compare the variance of the results of the runs with this factor at the high level to the variance of the runs with the lower values. In an ideal world, the ratio between these variances would be exactly one if the factor has no influence on the variance, and different from one otherwise. In the real world the ratio will never be exactly one, due to random variations. So, the question is how far from one can the ratio get before we decide that the factor has an impact on the variance (and can thus be used to reduce the variance).

Note that for this purpose 1/17, say, is exactly as meaningful as 17. By reversing the order of the numerator and denominator we can switch from 1/17 to 17 without any real change in the data. The F distribution is designed to tell whether the ratio is far enough from 1 to be deemed significant, but, for economy in tabulation, if requires that we transform values less than 1, such as 1/17, to their reciprocal values, such as 17.

So, to apply the F distribution analysis to check whether a particular factor (or interaction) impacts the variance, we simply take the sum of the squared residuals at the high level of the factor (or interaction), and divide it by the sum of squared residuals at the low level. If the result is less than 1 we replace it by its reciprocal. We than check whether the result is far enough from 1 to be considered significant. For an experiment with 16 runs we use the F table entry with 8 and 8 degrees of freedom for this purpose (since there are

8 residuals at the high level and 8 at the low). This usage is offered as a heuristic, assuming that after the regression we still retain all our original degrees of freedom. This is not strictly true, but makes no difference in terms of our ultimate usage below. When we apply the Daniel method the degrees of freedom have no direct role. Thus we proceed with this somewhat simplified presentation.

Alternatively, we can take the $S^2$ statistic at the high and low level and use 7 and 7 degrees of freedom instead. This would be justified if we doubt we neutralized all the real effects on the mean by the regression model. In such a case we do not make the implicit assumption that the mean residual is zero both at the high and low levels.

If we use the F distribution to check a small number of predetermined factors that were picked as candidates before the experiment, the results will be trustworthy. If we're going to try many factors and interactions, we face the problem of finding false significant effects simply by searching a large enough number.

This difficulty can be ameliorated by specifying higher significance levels. For instance, we might specify a 1% significance level instead of the traditional 5%.[3] Another approach is to use probability paper, as we did before to identify real effects (the Daniel method). The problem is to get hold of F probability paper. These are not generally available, however, because the F distribution depends on two parameters (degrees of freedom of the numerator and the denominator).

Some computer programs provide output that can be used to bypass this problem. They compute the p-value for each ratio tested. Suppose now that we tested 15 factors and interactions. Under the null hypothesis (that none of them impacts the variance) their p-values should behave like a sample from a uniform distribution between 0 and 1, i.e., U[0,1]. So, we can plot the p-values in a similar manner to the Daniel method (but we don't need a probability paper here, because the function of probability paper is to provide transformation to the uniform distribution, and this is not necessary for the uniform distribution itself).

---

[3]Incidentally, Fisher includes tables for his z distribution with the level of significance of 0.1% for which he gives credit to Deming. To quote Fisher [1990]: "Such high levels of significance are especially useful when the test we make is the most favourable out of a number which *a priori* might equally well have been chosen."

Alternately, let's investigate the general behavior of the log transformation applied to the ratio of sums of squares (as suggested above). Here, instead of making sure that F statistic exceeds 1 we take it's logarithm. Note now that if $F = 1$, which would be its median value if we don't invert it when necessary to cause it to exceed 1, the log will be zero, and if F is, say, 1/17, the log will simply be -log17, hence we also expect the transformation to be symmetric. Does that make it normal too? No. But it can still be approximated by the normal distribution, especially when the number of degrees of freedom is large or when the number of degrees of freedom in the nominator and denominator is equal or nearly equal [Fisher, 1990], as in our case. For a small number of degrees of freedom, the distribution is likely to have thicker tails than does the normal, so more points are likely to fall off the straight line, causing the potential retention of inert factors. But this may be considered a conservative approach in the exploration stage. Let us elaborate this point a bit more.

Our objective at the exploratory stage analysis is to discard as many inert factors and interactions as possible. This will leave us with less factors and interactions to handle in the search stage. But, we cannot discard factors and interactions that are inert in terms of impacting the mean response unless we have evidence that they are also inert in terms of variance. For instance, it is likely that the choice of the machine will not impact the average response as much as it will impact the variance; hence this factor might be discarded based on the initial search for significant effects, but should be retained if it's not inert in terms of impacting the variance.

If we use the Daniel method, with a regular normal distribution paper, for a distribution that has thicker tails, it's likely that some inert factors will appear significant, by mistake. This implies we will be less likely to discard such factors and interactions as inert. Thus we will tend to retain inert factors rather than reject significant ones. This gives us the chance to make corrections later, at the search stage. But factors that are discarded at the exploratory stage are lost forever. Thus, in conclusion, the usage of the Daniel method with the statistic proposed will be conservative in terms of not tending to discard significant effects, but tending to keep some insignificant ones (which can be discarded later).

## Distribution List

| Agency | No. of copies |
|---|---|
| Defense Technical Information Center<br>Cameron Station<br>Alexandria, VA 22314 | 2 |
| Dudley Knox Library, Code 0142<br>Naval Postgraduate School<br>Monterey, CA 93943 | 2 |
| Office of Research Administration<br>Code 08<br>Naval Postgraduate School<br>Monterey, CA 93943 | 1 |
| Library, Center for Naval Analyses<br>4401 Ford Avenue<br>Alexandria, VA 22302-0268 | 1 |
| Department of Administrative Sciences Library<br>Code AS<br>Naval Postgraduate School<br>Monterey, CA 93943 | 1 |
| Department of Administrative Sciences<br>Code AS/Tr<br>Naval Postgraduate School<br>Monterey, CA 93943 | 40 |